Integrating Data Uncertainty in Multiresolution Analysis

Marco S. Reis^{a,*}, Pedro M. Saraiva^a

^a GEPSI-PSE Group, Department of Chemical Engineering, University of Coimbra
Pólo II – Pinhal de Marrocos, 3030-290 COIMBRA, PORTUGAL

Abstract

Multiresolution decomposition frameworks are instrumental when one needs to focus data analysis at a particular scale or to separate the several contributions to the overall phenomenon, arising from different scales either in time or length. On the other hand, measurement uncertainties are now frequently available along with raw value measurements. In this paper we address the development of multiresolution decomposition methodologies that are able to cope with difficult data/uncertainty structures often met in industrial practice, like missing data and heteroscedastic uncertainties. A guideline is provided regarding their adequate use and an example illustrates the potential benefits associated with the suggested approach.

Keywords: multiresolution analysis, multiscale decomposition, wavelets, missing data, measurement uncertainty.

1. Introduction

Multiscale approaches have received considerable interest from several widely different branches of science, including PSE, where wavelet theory, and in particular Mallat's (1988) concept of multiresolution analysis have been playing an important role, particularly in what concerns data analysis applications (Mallat, 1998; Alsberg et al., 1997). Another theme that has gained considerable importance in the last years regards the specification and incorporation of uncertainty information in data analysis. Uncertainty is defined as a "parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand" (more details on ISO, 1993; Lira, 2002). With the appearance of new sensors, the development of metrology and the increasing enforcement driven by standardization organizations, one has now frequently available for analysis not only measurement values but also their uncertainties, i.e., two tables containing important information: one with raw data and another with their associated uncertainties.

In this paper we address the development of multiresolution decomposition methodologies that incorporate data uncertainty information, and are able to cope with difficult data/uncertainty structures that are often met in industrial practice, like missing data (i.e., when there are blanks in the raw data table due to, for instance, the existence of variables with different acquisition rates or problems either in the process or in the sensor/transmission/storage information chain) and heteroscedastic uncertainties

* Author to whom correspondence should be addressed: marco@eq.uc.pt

(signals noise with a variance that changes over time). In the next section we present two uncertainty-based multiresolution decomposition frameworks, and in section 3 we provide a general guideline for their use. Section 4 illustrates the potential benefits associated with the proposed approaches, and in section 5 we conclude by discussing some issues concerning the use of these approaches as well as other application contexts.

2. Accounting for Data Uncertainties in Multiresolution Decomposition (MRD)

Under the heading of multiresolution approximation, Mallat (1989) presented a framework where coarser approximations of a given signal at the finest scale can be considered as projections to approximation subspaces indexed by a scale index (j), V_j , that span progressively shorter regions of $L^2(\mathbb{R})$, and that have a nested structure ($V_{j+1} \subset V_j, V_{j+1} \neq V_j$). On the other hand, the details that are lost in the process of projecting the signal to increasingly coarser approximation spaces can also be considered as projections to complementary subspaces, the details spaces, W_j , that, in conjunction with the approximation space, span the space of the original signal. This allows us to write a given signal at the finest scale, say f_0 , as the sum of its projection to the approximation space at scale j , f_j , plus all the details relative to the scales in between, $\{w_i\}_{i=1,\dots,j}$. These projections can adequately be written in terms of linear combinations of the basis functions for the spaces, multiplied by the expansion coefficients. The expansion coefficients are called approximation coefficients, a_k^j ($k \in \mathbb{Z}$) (multipliers of the basis functions for the approximation space), and details coefficients, d_k^i ($i = 1, \dots, j; k \in \mathbb{Z}$) (multipliers of the basis functions for the detail spaces), and are usually referred to as the orthogonal wavelet transform or wavelet coefficients.

For the purposes of this paper, we will call “multiresolution decomposition framework” to the algorithm that provides the expansion coefficients, which, for the classical situation, is just the wavelet transform. However, this classical procedure can not be straightforwardly applied in less conventional situations, like those with missing data, something that occurs quite often in industrial scenarios, and furthermore does not take explicitly into account data uncertainties, when these are available. In the next subsections, we will present two methodologies that handle these issues, for situations with missing data and homoscedastic (constant) or heteroscedastic (varying) uncertainties (Method 1 and 2).

2.1 Method 1: Adjusting filter weights according to data uncertainties.

The Haar wavelet transform, perhaps the simplest of all known wavelet transform families, attributes a very clear meaning to its coefficients: approximation coefficients are averages of non-overlapping blocks with two elements (scaled by a factor of $1/\sqrt{2}$), and detail coefficients are calculated as the difference between the average and the first

element of the block. Cascading this procedure along the sets of approximation coefficients results in the Haar wavelet transform. Such a process gives equal weights to both values involved in the calculation of their average (coarser approximation coefficient). With data measurement uncertainties at our disposal, the averaging process can be modified so that the calculation of the coarser approximation coefficient ($a_{\lceil k/2 \rceil}^{j+1}$) gives more weight to the datum with less uncertainty. This can be achieved by using different averaging coefficients ($C_{\lceil k/2 \rceil}^{j+1,1}, C_{\lceil k/2 \rceil}^{j+1,2}$) applied to each datum (a_k^j, a_{k+1}^j):

$$a_{\lceil k/2 \rceil}^{j+1} = C_{\lceil k/2 \rceil}^{j+1,1} \cdot a_k^j + C_{\lceil k/2 \rceil}^{j+1,2} \cdot a_{k+1}^j \quad (1)$$

where $\lceil x \rceil$ is the smallest integer $n \geq x$.

Adequate weights can be set using formulas for the MVUE (minimum variance unbiased estimator) of the (common) average, suggesting the averaging coefficients associated to each datum presented in equation (2). Detail coefficients are given by (3).

$$C_{\lceil k/2 \rceil}^{j+1,1} = \frac{1/\text{unc}(a_k^j)^2}{1/\text{unc}(a_k^j)^2 + 1/\text{unc}(a_{k+1}^j)^2}, C_{\lceil k/2 \rceil}^{j+1,2} = 1 - C_{\lceil k/2 \rceil}^{j+1,1} \quad (2)$$

$$d_{\lceil k/2 \rceil}^{j+1} = C_{\lceil k/2 \rceil}^{j+1,1} \cdot (a_k^j - a_{\lceil k/2 \rceil}^{j+1}) = C_{\lceil k/2 \rceil}^{j+1,2} \cdot (a_{k+1}^j - a_{\lceil k/2 \rceil}^{j+1}) \quad (3)$$

where $\text{unc}(x)$ means the uncertainty associated with value x . Data uncertainty of the approximation coefficients at scale j should also be propagated to the approximation and detail coefficients at the next coarser scale $j+1$, so that the averaging procedure can continue to coarser scales, and to enable the specification of the uncertainties associated with each coefficient for posterior use. This can be done by applying the law of propagation of uncertainties (LPU) to the present situation (Lira, 2002):

$$\text{unc}(a_{\lceil k/2 \rceil}^{j+1}) = \sqrt{\left(C_{\lceil k/2 \rceil}^{j+1,1}\right)^2 \cdot \text{unc}(a_k^j)^2 + \left(C_{\lceil k/2 \rceil}^{j+1,2}\right)^2 \cdot \text{unc}(a_{k+1}^j)^2} \quad (4)$$

$$\text{unc}(d_{\lceil k/2 \rceil}^{j+1}) = \sqrt{\left(C_{\lceil k/2 \rceil}^{j+1,1}\right)^2 \cdot \left(\text{unc}(a_k^j)^2 + \text{unc}(a_{k+1}^j)^2 - 2 \cdot C_{\lceil k/2 \rceil}^{j+1,1} \cdot \text{unc}(a_k^j)\right)} \quad (5)$$

where it is assumed that the errors from two successive observations are independent, but more complex situations can also be considered. By conducting a multiresolution decomposition using this procedure, we give more weight to the values with less uncertainty associated during the calculation of the approximation coefficients. Extending this reasoning further, we can see that this calculation scheme allows for the integration of missing data in the analysis, as a missing datum can be considered to be any finite number with an infinite uncertainty associated, that essentially removes it from (2)-(5). In this case, the coarser approximation coefficient will have the same value as the non-missing datum and the coarser detail coefficient will be zero. For the case where there is no missing data and data uncertainties are homoscedastic, this multiresolution decomposition framework will provide the same values as the Haar transform (up to a scaling factor of $2^{j/2}$ for the coefficients of scale j).

2.2 Method 2: Use Haar wavelet filter, accommodate missing data and propagate data uncertainties to coarser coefficients

In this second approach for incorporating data uncertainties, the averaging and differencing coefficients are kept constant and equal to the ones suggested by the Haar filter. When there are no data missing, the uncertainty of the finer approximation coefficients is propagated to the coarser approximation and detail coefficients, using the law of propagation of uncertainties. If there are missing data, we calculate the next coarser coefficients by successively applying the following rules to each new pair of approximation coefficients at scale j , $\{a_k^j, a_{k+1}^j\}$:

- No value is missing $\Rightarrow$ use Haar and calculate uncertainties using LPU;
- $\{a_k^j\}$ is missing $\Rightarrow a_{\lfloor k/2 \rfloor}^{j+1} = a_{k+1}^j, unc(a_{\lfloor k/2 \rfloor}^{j+1}) = unc(a_{k+1}^j), d_{\lfloor k/2 \rfloor}^{j+1} = 0, unc(d_{\lfloor k/2 \rfloor}^{j+1}) = 0$;
- $\{a_{k+1}^j\}$ is missing $\Rightarrow a_{\lfloor k/2 \rfloor}^{j+1} = a_k^j, unc(a_{\lfloor k/2 \rfloor}^{j+1}) = unc(a_k^j), d_{\lfloor k/2 \rfloor}^{j+1} = 0, unc(d_{\lfloor k/2 \rfloor}^{j+1}) = 0$;
- $\{a_k^j, a_{k+1}^j\}$ are missing $\Rightarrow a_{\lfloor k/2 \rfloor}^{j+1} = \text{missing}, unc(a_{\lfloor k/2 \rfloor}^{j+1}) = \text{missing};$
 $d_{\lfloor k/2 \rfloor}^{j+1} = \text{missing}, unc(d_{\lfloor k/2 \rfloor}^{j+1}) = \text{missing}.$

From the rules above, we can see that when no missing data is present, the procedure consists of applying the Haar wavelet with uncertainty propagation. When we have missing data, it can also happen that it remains present at coarser scales (see the fourth rule). This can be instrumental when analysing the information content at different scales, and enables the development of tools for scale selection.

3. Guidelines on the Use of Uncertainty-Based MRD Frameworks

Methods 1 and 2 differ deeply on how they implement the incorporation of uncertainty information. In this section we provide a general guideline about which of the two approaches to use and when. Let us consider an artificial piecewise constant signal, where the values are held constant in windows of $2^4 = 16$ successive values (Figure 1-a), to which proportional noise with uncertainty assumedly known is added. Using the noisy signal (Figure 1-b), it is possible to calculate its approximations for coarser scales ($j = 1, 2, \dots$) using the two types of approaches and then to see which method performs better in the task of approximating the true signal when projected at the same scale, j . Our performance index is the mean square error between the approximation at scale j , calculated for the noisy signal, and that for the true signal, $MSE(j)$. Figure 1-c summarizes the results obtained from 100 simulations. These results illustrate the general guideline according to which, from the strict stand point of the approximation ability at coarser scales, Method 1 is more adequate than Method 2 for constant signals and for piecewise constant signals until we reach the scale where the true values begin to vary from observation to observation, i.e., for which the piecewise constant behaviour stops. As the original signal has constant values along windows of 16 values, the piecewise constant pattern breaks down after scale $j = 4$. This occurs because Method 1 is based on the MVUE estimator of an underlying constant mean for two successive values, thus leading to improved results when this assumption holds, at least

approximately as in the case of piecewise constant signals, being overtaken by Method 2 when such an assumption is no longer valid.

Figure 1. (a) true signal used in the simulation; (b) a realization of the noisy signal and (c) box plots for the difference in MSE at each scale (j) obtained for the two methods 100 simulations).

4. An Uncertainty-Based De-Noising Application

Wavelets found great success in the task of “cleaning signals” from undesirable components of stochastic nature, called noise. If we are in such a position that we do know the main noise features, namely measurement uncertainties, then we can use this additional piece of information to come up with simple but effective de-noising schemes. As an illustration, consider a smoothed version of a NIR spectrum as the “true” signal, to which heteroscedastic proportional noise is added. The standard de-noising procedure was then applied to the noisy signal, according to the following sequence of steps: 1. Decomposition of the signal into its wavelet coefficients; 2. Application of a thresholding technique to the calculated coefficients; 3. Reconstruction of the signal using processed coefficients. This procedure is tested for the classic Haar wavelet with the threshold suggested by Donoho and Johnstone (1992), $T = \hat{\sigma}\sqrt{2\ln(N)}$, where $\hat{\sigma}$ is a robust estimator of the noise (constant) standard deviation, along with a “Translation Invariant” extension of it, based on Coifman’s “Cycle Spinning” concept (Coifman and Donoho, 1995): “Average[Shift – De-noise – Unshift]”, where all possible shifts were used. We will call this alternative as “TI Haar”. These methods are to be compared with their counterpart procedures that have the ability of using explicitly the available uncertainty information: “Haar + uncertainty propagation”, “TI Haar + uncertainty propagation” (only 10 rotations were used in this method). All the tested methods used the same wavelet (Haar), threshold constant

($\sqrt{2\ln(N)}$) and thresholding policy (“Hard Threshold”). Figure 2 presents results for the MSE scores of the reconstructed signal (scale $j = 0$) relatively to the true one, obtained for 100 realizations of the additive noise. A clear improvement in MSE is found for the uncertainty-based methods, relatively to their classic counterparts.

Figure 2. De-noising results associated with the four alternative methodologies (“Haar”, “Haar+uncertainty propagation”, “TI Haar” and “TI Haar+uncertainty propagation”) for 100 noise realizations.

5. Conclusions

In this paper, we propose two methods for handling the issues of missing data and data uncertainty in MRD. Both Methods 1 and 2 are not extensions of the wavelet transform in a strict sense, as some of their fundamental properties do not always hold, such as the energy conservation property (in the sense of the Plancherel formula; Mallat, 1998). However, they can lead to improved results by effectively incorporating uncertainty information and allow one to extend the wavelet MRD to contexts where it could not be directly applied, namely when we have missing data, as well as provide new tools for addressing other types of problems in data analysis, such as the one of selecting a proper scale for data analysis.

References

- Alsberg, B.K., A.M. Woodward, D.B. Kell, 1997, Chemometrics Intell. Lab. Syst. 37, p. 215-239.
- ISO, 1993, Guide to the Expression of Uncertainty. Geneva, Switzerland.
- Lira, I., 2002, Evaluating the Measurement Uncertainty, NJ: Institute of Physics Publishing.
- Mallat, S., 1989, IEEE Trans. on Pattern Analysis and Machine Intell. 11, 7, p. 674-693.
- Mallat, S., 1998, A Wavelet Tour of Signal Processing. San Diego [etc.]: Academic Press.

Acknowledgements

The authors would like to acknowledge Portuguese FCT for financial support through research project POCTI/EQU/47638/2002.